Add Health

The National Longitudinal Study of Adolescent to Adult Health

How to Obtain Add Health *OMICs* Data

02/11/19

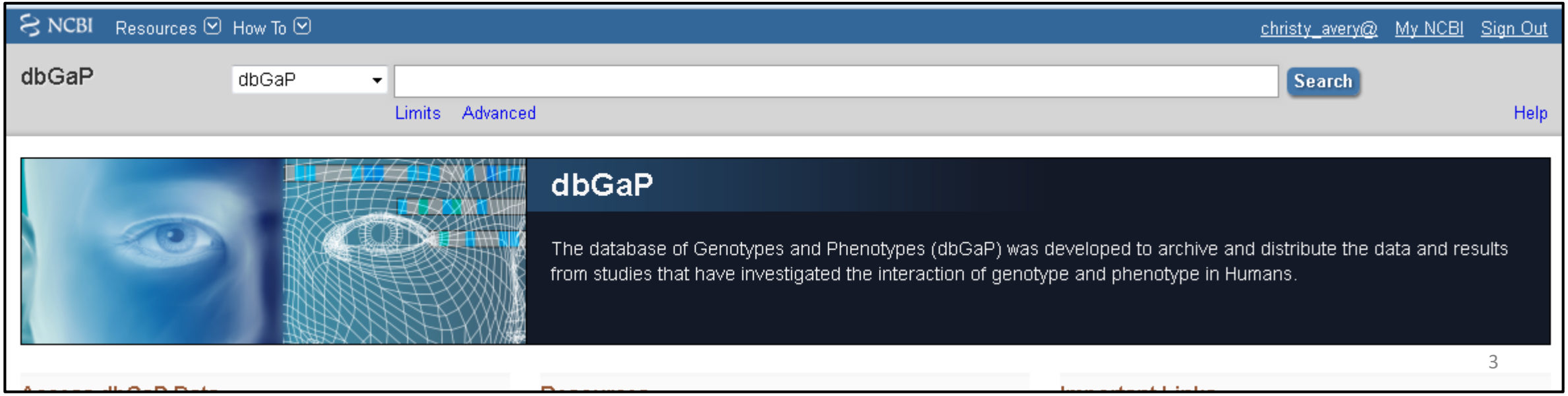
addhealth_genetics@unc.edu

Overview: How to Access Add Health GWAS Data

- The goal of NIH/NICHD-funded grant R03HD097630 (MPIs: Mullan Harris, Avery) is to develop research tools to enable widespread access and use of Add Health genomics (e.g. GWAS, exome etc.) data. This slide set provides tailored instruction in how to access Add Health genomics data.
- ***Note***: Add Health phenotype data are available through: https://www.cpc.unc.edu/projects/addhealth/contracts.
  - See accompanying slide set on accessing additional phenotype data.

dbGaP: NIH Genomics Warehouse

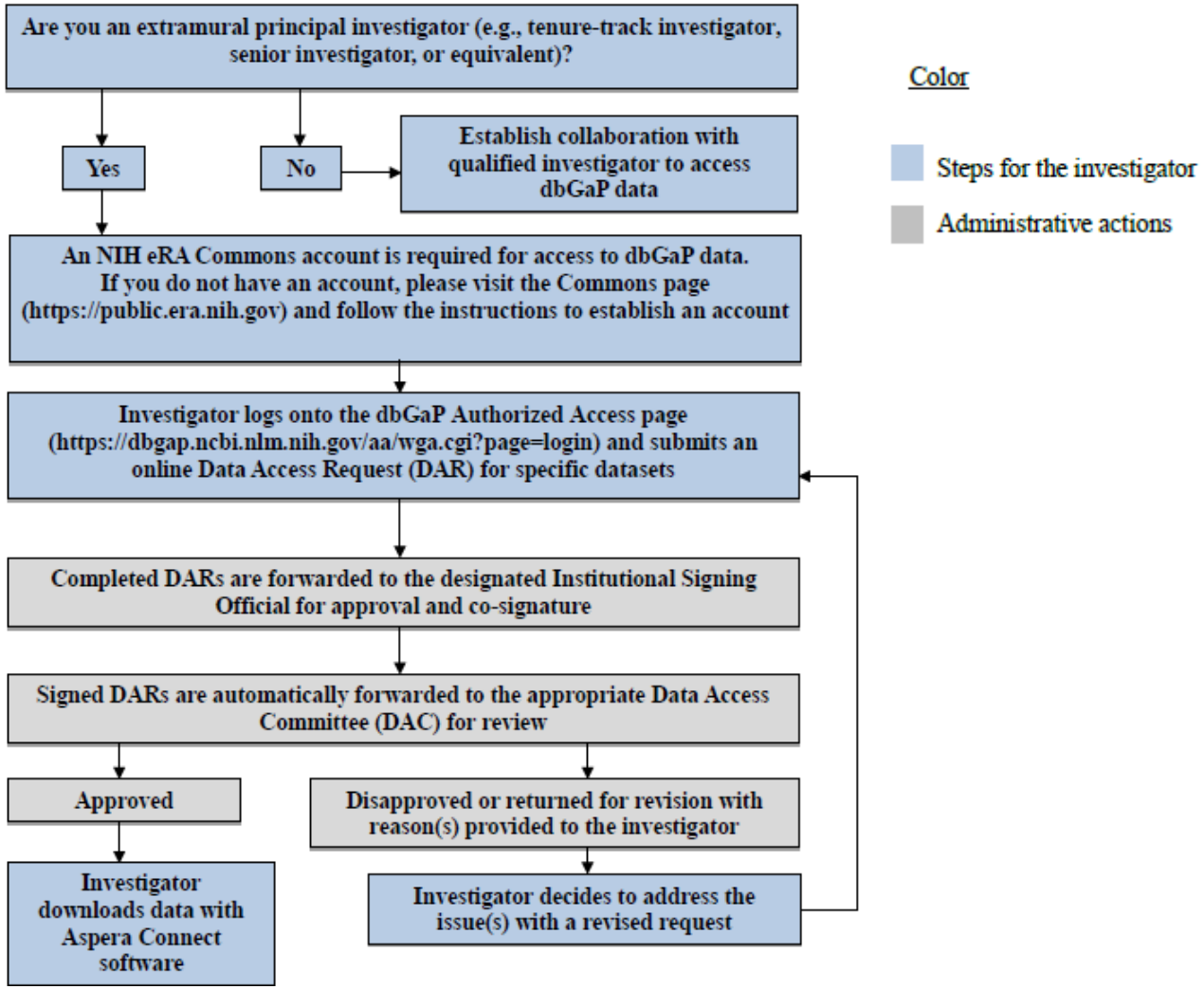
- Add Health GWAS data and accompanying documentation are available from the NIH-sanctioned database of Genotypes and Phenotypes (dbGaP), a repository for archiving, curating, and distributing GWAS data. https://www.ncbi.nlm.nih.gov/gap



Who can Apply for Data Through dbGaP?

- **Non-NIH investigators (i.e. Extramural investigators)**
  - Extramural Investigators must be permanent employees of their institution at a level equivalent to a tenure-track professor or senior scientist with responsibilities that most likely include laboratory administration and oversight. Laboratory staff and trainees such as graduate students and postdoctoral fellows are not permitted to submit dbGaP project requests.
- **NIH Investigators:**
  - NIH Intramural Investigators must be tenure-track investigators, senior investigators, senior scientists, senior clinicians, or staff scientists.
  - NIH extramural scientific staff must have administrative responsibility for the data; have substantial research involvement in the award that generated the data; or need access to carry out research unrelated to their portfolio management responsibilities.

Overview of Process Investigators to Access Datasets in dbGaP



https://osp.od.nih.gov/wp-content/uploads/Extramural_Flowcharts_for_Data_Access.pdf

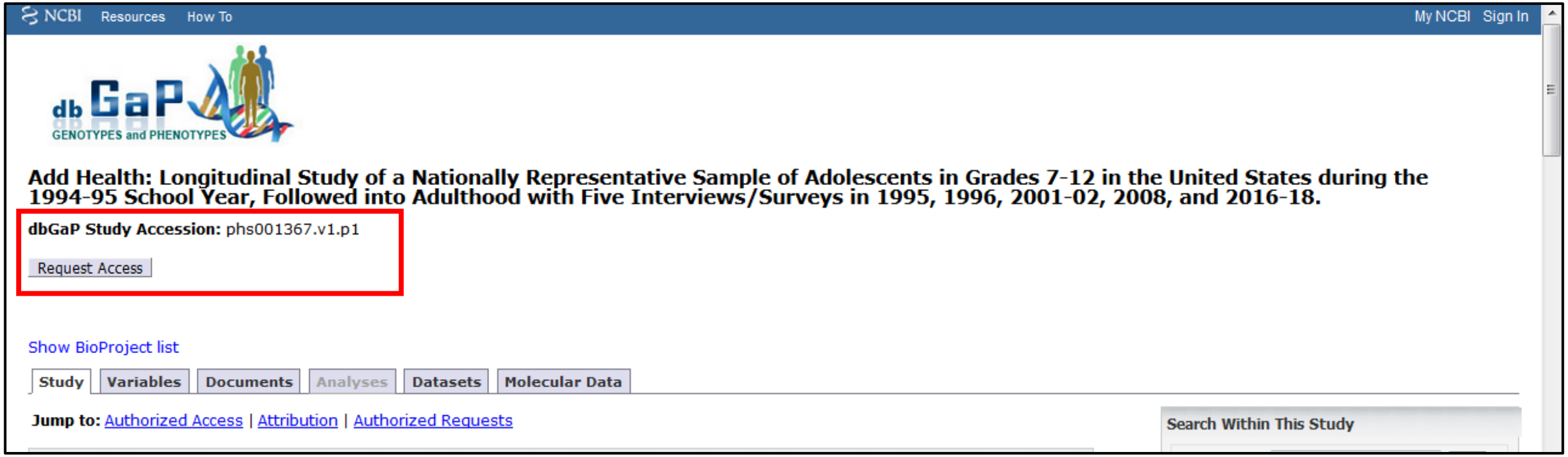
First Step: Obtain an eRA Commons Account

- To log into dbGap and request access to controlled-access datasets, you must have an **eRA Commons** account (https://era.nih.gov/commons-account-information.cfm).
- If you do not have a pre-existing account, register here: https://era.nih.gov/reg_accounts/register_commons.cfm.

Electronic Research Administration
A program of the National Institutes of Health

Identify Add Health GWAS Data and Request Access

- Add Health study accession number: phs001367.v1.p1



https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001367.v1.p1

A Completed dbGaP Application:

- Is identified by a Data Access Request (DAR) number and a project number
- Is project-specific. Approval for one project does not carry over to a new project.
- Requires the applicant to review and agree to terms, conditions, and statements of the Add Health Data Use Certification Agreement (October 27, 2015 version, which is currently being updated).

Information Needed for dbGaP Application

- Research statement and nontechnical summary statement describing your planned use of the data specific to your project
- Name of the institutional signing official
- A list of internal investigators at your institution who will share access to the data
- A list of external collaborating investigators
- The name of the information technology (IT) director
- Local Institutional Review Board (IRB) approval.

dbGaP Approved User Code of Conduct

- Investigator(s) will use requested datasets solely in connection with the research project described in the approved Data Access Request for each dataset;
- Investigator(s) will make no attempt to identify or contact individual participants from whom these data were collected without appropriate approvals from the relevant IRBs;
- Investigator(s) will not distribute these data to any entity or individual beyond those specified in the approved Data Access Request;
- Investigator(s) will adhere to computer security practices that ensure that only authorized individuals can gain access to data files;
- Investigator(s) will not submit for publication or any other form of public dissemination analyses or other reports on work using or referencing NIH datasets prior to the embargo release date listed for the dataset (or dataset version) on dbGaP;
- Investigator(s) acknowledge the Intellectual Property Policies as specified in the Data Use Certification; and,
- Investigator(s) will report any inadvertent data release in accordance with the terms in the Data Use Certification, breach of data security, or other data management incidents contrary to the terms of data access

Begin New Research Project

NCBI | Site map | All databases | PubMed | Search

db GaP genotypes and phenotypes | Browse/Search | Authorized Access | Help

Logged in as **Christy Avery** | Log out

Beacon | Data Browser | My Projects | My Requests | Downloads | Downloaders | My Profile

My Research Project

General Instructions

- This application will automatically generate a Data Access Request (DAR) number and a project number. Please keep track of this number for future communications with dbGaP and relevant Data Access Committee(s) (DAC)
- A completed request for data access includes this form as well as a review of and agreement to the terms, conditions, and statements in the Data Use Certification (DUC) for each respective dataset requested.
- Dataset requests are project-specific. If you were granted access to a dataset(s) for another project, that approval does not carry over to this new proposed project. You must request access to all datasets that you plan to use in the new project.
- Please note that fields marked as "*" are required fields.

Before You Get Started

In order to complete the application for data access you will need to collect the following information:

- A research statement and a nontechnical summary statement describing your planned use of the data.
- The name of the institutional signing official who will certify the terms of use assurances on behalf of your institution.
- A list of all internal investigators at your institution who will share access to the data for the proposed research.
- A list of external collaborating investigators.
- The name of the information technology (IT) Director.
- Some datasets may require local Institutional Review Board (IRB) approval for use. These are noted in the study list. Please check the individual study pages in dbGaP for these additional requirements.
- Some datasets may require supplemental documentation to accompany this standard application. Review the DUC* instruction pages for detailed information about how to prepare these materials in a single PDF file.

* You can navigate to each study DUC from the public study home page in dbGaP. Look for the "individual-level data" section.

dbGaP APPROVED USER CODE OF CONDUCT

The following is the Code of Conduct that research investigators agree to abide by as Approved Users of data received through the database of Genotypes and Phenotypes (dbGaP). Failure to abide by any term within this Code of Conduct may result in revocation of approved access to any or all datasets obtained through dbGaP.

The elements of the NIH Code of Conduct for Data Use include:

1. Investigator(s) will use requested datasets solely in connection with the research project described in the approved Data Access Request for each dataset;
2. Investigator(s) will make no attempt to identify or contact individual participants from whom these data were collected without appropriate approvals from the relevant IRBs;
3. Investigator(s) will not distribute these data to any entity or individual beyond those specified in the approved Data Access Request;
4. Investigator(s) will adhere to computer security practices that ensure that only authorized individuals can gain access to data files;
5. Investigator(s) will not submit for publication or any other form of public dissemination analyses or other reports on work using or referencing NIH datasets prior to the embargo release date listed for the dataset (or dataset version) on dbGaP;
6. Investigator(s) acknowledge the Intellectual Property Policies as specified in the Data Use Certification; and,
7. Investigator(s) will report any inadvertent data release in accordance with the terms in the Data Use Certification, breach of data security, or other data management incidents contrary to the terms of data access.

Begin New Research Project

NIH Genotype and Phenotype database is a service of NCBI. Please **contact us** with any questions.

National Center for Biotechnology Information | U.S. National Library of Medicine

Privacy Notice | Disclaimer | Accessibility

Identify Add Health Study (phs001367.v1.p1)

dbGaP: Authorized Access: Pro

https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?page=newprj_choosedatasets

Updates HeelMail epid_is clock phone weather bus Google

NCBI Site map | All databases | PubMed | Search

db GaP genotypes and phenotypes | Browse/Search | Authorized Access | Help

Logged in as **Christy Avery** | Log out

Beacon | Data Browser | My Projects | My Requests | Downloads | Downloaders | My Profile

Project Request

+ OMB control number: 0925-0670 Expiration date: 03/31/2019

Research Project | Cloud Providers | Collaborators | IT Director | Confirm Datasets | Review DUC | Review DUL | Review Applications | Feedback

Please select datasets to request access to. If you have changed any common information (research statement, list of collaborators), all approved application and those being reviewed by DAC(s) will need to be resubmitted.
For any study that has more than one consent group, there are no overlaps in subjects between the consent groups.

Filter Consents | Clear Filter

Primary disease | Molecular data type | Study design

Approved for GRU | Approved for commercial use | Approved for method development | Health biomedical research

Study lookup enter study accession | Study accession | Exclude IRB required

phs001363 - The 100-Person Wellness Project (HPWP)
phs001364 - Mapping the Human Connectome: Structure, Function, and Heritability
phs001365 - Neoadjuvant Chemotherapy +/- Bevacizumab Pharmacogenomics
phs001366 - Pilot Study of Exome and mRNA Profiling in Small Cell Lung Cancer Patients
phs001367 - Add Health: The National Longitudinal Study of Adolescent to Adult Health (Add Health)
phs001368 - NHLBI TOPMed: Cardiovascular Health Study
phs001371 - TNF-induced dynamic regulation of mRNA stabilome in rheumatoid arthritis fibroblast-like synoviocytes
phs001373 - Mechanisms Controlling Human Microglia Gene Expression
phs001374 - VA APOLLO Project - Research for Precision Oncology (RePOP)
phs001376 - CD55 Deficiency, Early-Onset Protein-Losing Enteropathy and Thrombosis
phs001377 - Epigenetics of Cocaine and Nicotine Addiction
phs001381 - Gut Microbiome and Types of Colorectal Polyps
phs001382 - Reproductive Health in Men and Women with Vasculitis - VCRC 5531
phs001384 - Time Lapse to Cancer-Defining the transistion from polyp to cancer
phs001385 - Ancestry Admixture among Chileans
phs001386 - Smoking and the Vaginal Microbiome
phs001389 - Gene Expression and Histone Modification Profiles in GBM Tumors
phs001391 - OncoArray: Prostate Cancer
phs001392 - A Prospective Natural History Study of Diagnosis, Treatment and Outcomes of Children with SCID Disorders - PIDTC 6901
phs001394 - Hypertension-Insulin Resistance Family Study (HTN-IR)

Page 1 2 3 4 5 6 7 8 ... 21

Consent Group	Participants	DAR Status
Genetic Epidemiology of Age-Related Macular Degeneration in the Older C		
General Research Use (IRB) (phs001361.v1.p1.c1), NEI		
Glaucoma Exome Sequencing (phs000558.v1.p1)		
General Research Use (IRB) (phs000558.v1.p1.c1), NEI		
The 100-Person Wellness Project (HPWP) (phs001363.v1.p1)		
General Research Use (IRB, PUB, COL, NPU) (phs001363.v1.p1.c1), NIGMS		
The Genetic Landscape of Metastasis and Recurrence in HNSCC (phs0010		
Disease-Specific (Cancer, MDS) (phs001007.v1.p1.c1), NCI DAC		
Up for a Challenge: African American Breast Cancer Consortium (AABC) S		
Up for a Challenge (Publication required) (phs000851.v1.p1.c1), NCI DAC		
Up for a Challenge (Not for Profit Use Only, Publication required)	Use of this data is limited to research described for the National Cancer Institute (NCI) "Up for A Challenge" breast cancer genetic epidemiology competition. The goal of	486

Accessing New OMICs Files

- Note that the study accession number, phs001367.**v1**.p1, indexes the version
- New versions are created when new data are uploaded (e.g. exome chip data per R01HD057194). Add Health users with an active dbGaP contract will be alerted to new releases and granted access to newly available data.

dbGaP: Authorized Access: Proj

https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?page=newprj_choosedatasets&filter=wlid&wlid=&filter=fc_gru&filter=fc_npu&filter=

Updates | HeelMail | epid_is | clock | phone | weather | bus | Google

NCBI | Site map | All databases | PubMed | Search

db GaP genotypes and phenotypes | Browse/Search | Authorized Access | Help

Logged in as **Christy Avery** | Log out

Beacon | Data Browser | My Projects | My Requests | Downloads | Downloaders | My Profile

Project Request

+ OMB control number: 0925-0670 Expiration date: 03/31/2019

Research Project | Cloud Providers | Collaborators | IT Director | Confirm Datasets | Review DUC | Review DUL | Review Applications | Feedback

Please select datasets to request access to. If you have changed any common information (research statement, list of collaborators), all approved application and those being reviewed by DAC(s) will need to be resubmitted.
For any study that has more than one consent group, there are no overlaps in subjects between the consent groups.

Filter Consents [Clear Filter]

Primary disease | Molecular data type | Study design

☐ Approved for GRU ☐ Approved for commercial use ☐ Approved for method development ☐ Health biomedical research

Study lookup: enter study accession | Study accession: phs001367 - Add Health: The National Longitudinal Study of . | ☐ Exclude IRB required

Consent Group	Data Use Limitations	Participants	DAR Status
Add Health: The National Longitudinal Study of Adolescent to Adult Health (Add Health) (phs001367.v1.p1)			
☑ General Research Use (IRB, PUB, GSO) (phs001367.v1.p1.c1), NICHD	Use of the data is limited only by the terms of the model Data Use Certification. Requestor must provide documentation of local IRB approval. Requestor agrees to make results of studies using the data available to the larger scientific community. Use of the data is limited to genetic studies only. . **This consent group requires IRB approval attachment**	9974	

[Return to My Projects] [Add Selected and Continue]

Study accession for preview: [ ] [Add]

This input box is only for study investigators of studies that are currently in preview status. If you are a data submitter, please input the study accession.

NIH Genotype and Phenotype database is a service of NCBI. Please **contact us** with any questions.

National Center for Biotechnology Information | U.S. National Library of Medicine

Privacy Notice | Disclaimer | Accessibility

NIH | FIRSTGOV.gov

Prepare and Enter/Identify Your Title, Research Use Statement, Summary, and Signing Official

dbGaP: Authorized Access: Proj

https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?

Updates HeelMail epid_is clock phone weather bus Google

NCBI Site map | All databases | PubMed | Search

db GaP genotypes and phenotypes | Browse/Search | Authorized Access | Help

Logged in as **Christy Avery** | Log out

Beacon | Data Browser | My Projects | My Requests | Downloads | Downloaders | My Profile

Project Request

#20018:

SO:

+ OMB control number: 0925-0670 Expiration date: 03/31/2019

Project Details | Choose Datasets | Research Project | Cloud Providers | Collaborators | IT Director | Confirm Datasets | Review DUC | Review DUL | Review Applications | Feedback

***Descriptive Title of Project**
Please note that coordinated requests by collaborating institutions should each use the same title.

Copy and paste your Research Use Statement and non-technical summary below. All applications must be made in English.

***Research Use Statement (RUS)**
A RUS is a brief description of the applicant's proposed use of dbGaP dataset(s). The RUS will be reviewed by all NIH Institutes and Centers responsible for data covered by this Data Access Request. Please note that if access is approved, you agree that the RUS, along with your name and institution, will be included on the dbGaP website to describe your research project to the public.

Please make it clear whether you plan to combine requested datasets with other datasets outside of dbGaP, and, if so, whether you plan to analyze these datasets independently or together. If you do plan to combine datasets in any way, please describe your plan and also please discuss whether it creates any additional risks to participants. If you are focusing on outcomes or hypotheses that were not the focus of the primary study (or studies), please describe the outcomes you propose to examine.

Investigators do not need to submit a new project request unless the dataset will be used for research outside of the scope of the approved Research Use Statement

Please enter your RUS in the area below. The RUS should be one or two paragraphs in length and include research objectives, the study design, and an analysis plan (including the phenotypic characteristics that will be tested for association with genetic variants). If you are requesting multiple datasets, please describe how you will use them. Examples of RUS can be found at GDS website. Please limit your RUS to 4500 characters.

☐ **I am requesting permission to use cloud computing to carry out the research as described in my Research Use Statement.**

***Non-technical summary**
Please enter below a non-technical summary of your RUS suitable for understanding by the general public (written at a high school reading level or below). Please limit your non-technical summary to 1300 characters.

***Choose your Signing Official (SO):**
Your SO is typically the same person who signs your grant applications and is an individual listed in eRA Commons as a SO for your institution and who has the authority to certify your application on behalf of your institution.

Cloud Computing

***Research Use Statement (RUS)**

A RUS is a brief description of the applicant's proposed use of dbGaP dataset(s). The RUS will be reviewed by all NIH Institutes and Centers responsible for data covered by this Data Access Request. Please note that if access is approved, you agree that the RUS, along with your name and institution, will be included on the dbGaP website to describe your research project to the public.

Please make it clear whether you plan to combine requested datasets with other datasets outside of dbGaP, and, if so, whether you plan to analyze these datasets independently or together. If you do plan to combine datasets in any way, please describe your plan and also please discuss whether it creates any additional risks to participants. If you are focusing on outcomes or hypotheses that were not the focus of the primary study (or studies), please describe the outcomes you propose to examine.

Investigators do not need to submit a new project request unless the dataset will be used for research outside of the scope of the approved Research Use Statement

Please enter your RUS in the area below. The RUS should be one or two paragraphs in length and include research objectives, the study design, and an analysis plan (including the phenotypic characteristics that will be tested for association with genetic variants). If you are requesting multiple datasets, please describe how you will use them. Examples of RUS can be found at GDS website. Please limit your RUS to 4500 characters.

☐ **I am requesting permission to use cloud computing to carry out the research as described in my Research Use Statement.**

- Add Health investigators are currently investigating the feasibility of providing users with virtual machine templates that meet required security protocols when using cloud computing.
- Until these security templates are available, required data security standards for remote compute servers are available here (see section for compute server, not file server):
  - https://www.cpc.unc.edu/research/tools/datasecurity/how-to-secure-a-server
- Outside of the security templates, the Add Health study cannot support costs associated with cloud data storage or analysis.

Do Not Forget Your Decryption Password You Will Need it When You Retrieve the Repository Key!

Create Decryption Password.
The files distributed through the dbGaP system are encrypted. A password is required for decrypting downloaded files. Please provide a decryption password for the project.
Valid passwords must be at least 8 ASCII characters long and must contain at least 3 of the following 4 characters:

- upper case letters
- lower case letters
- numbers
- non-alphanumeric characters

*Password for project: ••••••••••••

*Password confirmation: ••••••••••••

Back | Return to My Projects | Save | Save and Continue

NIH Genotype and Phenotype database is a service of NCBI. Please **contact us** with any questions.

Have Your IT Director and Collaborators Identified, Including Contact Information

dbGaP: Authorized Access: Proj

https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?

Updates | HeelMail | epid_is | clock | phone | weather | bus | Google

NCBI | Site map | All databases | PubMed | Search

db GaP genotypes and phenotypes | Browse/Search | Authorized Access | Help

Logged in as **Christy Avery** | Log out

Beacon | Data Browser | My Projects | My Requests | Downloads | Downloaders | My Profile

Project Request

#20018: asdfasdf

SO: Jenifer Gwaltney

+ OMB control number: 0925-0670 Expiration date: 03/31/2019

Project Details | Choose Datasets | Research Project | Collaborators | IT Director | Confirm Datasets | Review DUC | Review DUL | Review Applications | Feedback

An information technology (IT) director's (or designee's) contact information is required to ensure data security policies and procedures are in place. This individual must have the authority to vouch for the IT capabilities at your institution.

IT Director

Prefix	*First name	Middle name	*Last name	Suffix
	Donald		Draper	

*Position/Title	Department	*Organization name	Division
IT Director		UNIV OF NORTH CAROLINA CHAPEL HIL	

*Street1	Street2	*City	State	*ZIP/Postal code	*Country
123 W. Franklin St.		Chapel Hill	NC	27516	US

*E-mail	*Phone	Fax
ddd@unc.edu	919	

Back | Return to My Projects | Save | Save and Continue

NIH Genotype and Phenotype database is a service of NCBI. Please **contact us** with any questions.

National Center for Biotechnology Information | U.S. National Library of Medicine

Privacy Notice | Disclaimer | Accessibility

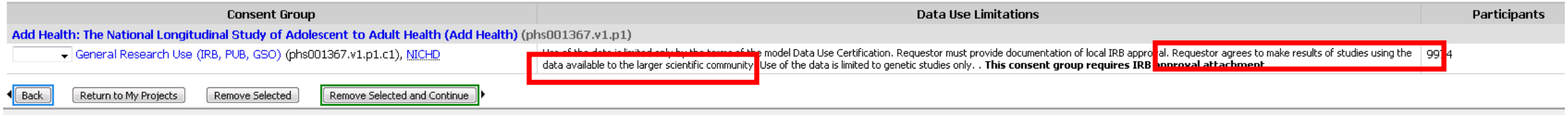
Add Heath Requires IRB Approval Prior to dbGaP Submission

Consent Group	Data Use Limitations	Participants
Add Health: The National Longitudinal Study of Adolescent to Adult Health (Add Health) (phs001367.v1.p1)		
General Research Use (IRB, PUB, GSO) (phs001367.v1.p1.c1), NICHD	Use of the data is limited only by the terms of the model Data Use Certification. Requestor must pr[illegible] make results of studies using the data available to the larger scientific community. Use of the data is limited to genetic studies only. **This consent group requires IRB approval attachment**	9974

Back | Return to My Projects | Remove Selected | Remove Selected and Continue

- Either expedited or exempt IRB approval is acceptable.

Publication of Genomic Summary Results



- Given the sensitivity of Add Health, genomic summary results (GSR) that contain Add Health data should be provided only through controlled-access data access request and review procedures (e.g. through dbGaP).
- For more information, see: https://grants.nih.gov/grants/guide/notice-files/NOT-OD-19-023.html

Submit Application!

dbGaP: Authorized Access: Proj

https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?

Updates | HeelMail | epid_is | clock | phone | weather | bus | Google

NCBI | Site map | All databases | PubMed | Search

db GaP genotypes and phenotypes | Browse/Search | **Authorized Access** | Help

Logged in as **Christy Avery** | Log out

Beacon | Data Browser | **My Projects** | My Requests | Downloads | Downloaders | My Profile

Project Request

#20018: asdfasdf

SO: Jenifer Gwaltney

+ OMB control number: 0925-0670 Expiration date: 03/31/2019

Project Details | Choose Datasets | Research Project | Collaborators | IT Director | Confirm Datasets | Review DUC | Review DUL | Review Applications | Feedback

Review and submit data access requests

The following application is the official request document that will be sent to your signing official (SO). Please note that you **will not be allowed** to change your application while it is being reviewed by the SO. In order to make the changes after you have submitted your application for review you will have to contact your SO with a request to return it for your revision.

After approval by your SO, each application will be sent to the appropriate Data Access Committee (DAC). Multiple DACs may need to evaluate your application.

Review Complete Application

Check the "I agree" boxes to provide the required certifications and assurances.

By signing below, I certify that the statements herein are true, complete, and accurate to the best of my knowledge. I am aware that any false, fictitious, or fraudulent statements or claims may subject me to criminal, civil, or administrative penalties.

☐ I agree

By signing below, I certify that I have read and agreed to the terms, conditions, and statements in the Data Use Certification(s) for the request dataset(s). I agree to abide by the Code of Conduct.

☐ I agree

Submit Application To Signing Official

Please **carefully review** submitted information by clicking on "Review Complete Application" .

To revise: Use navigation tabs to return to any previous step and make the changes.

To approve: **Check the boxes** below to attach your electronic signature acknowledging agreement to the terms.

You need Adobe Reader installed to review PDF forms.

Get ADOBE® READER®

This project currently contains **1 active request** for data access. You can view individual applications and processing statuses in the table below.

Active (1)

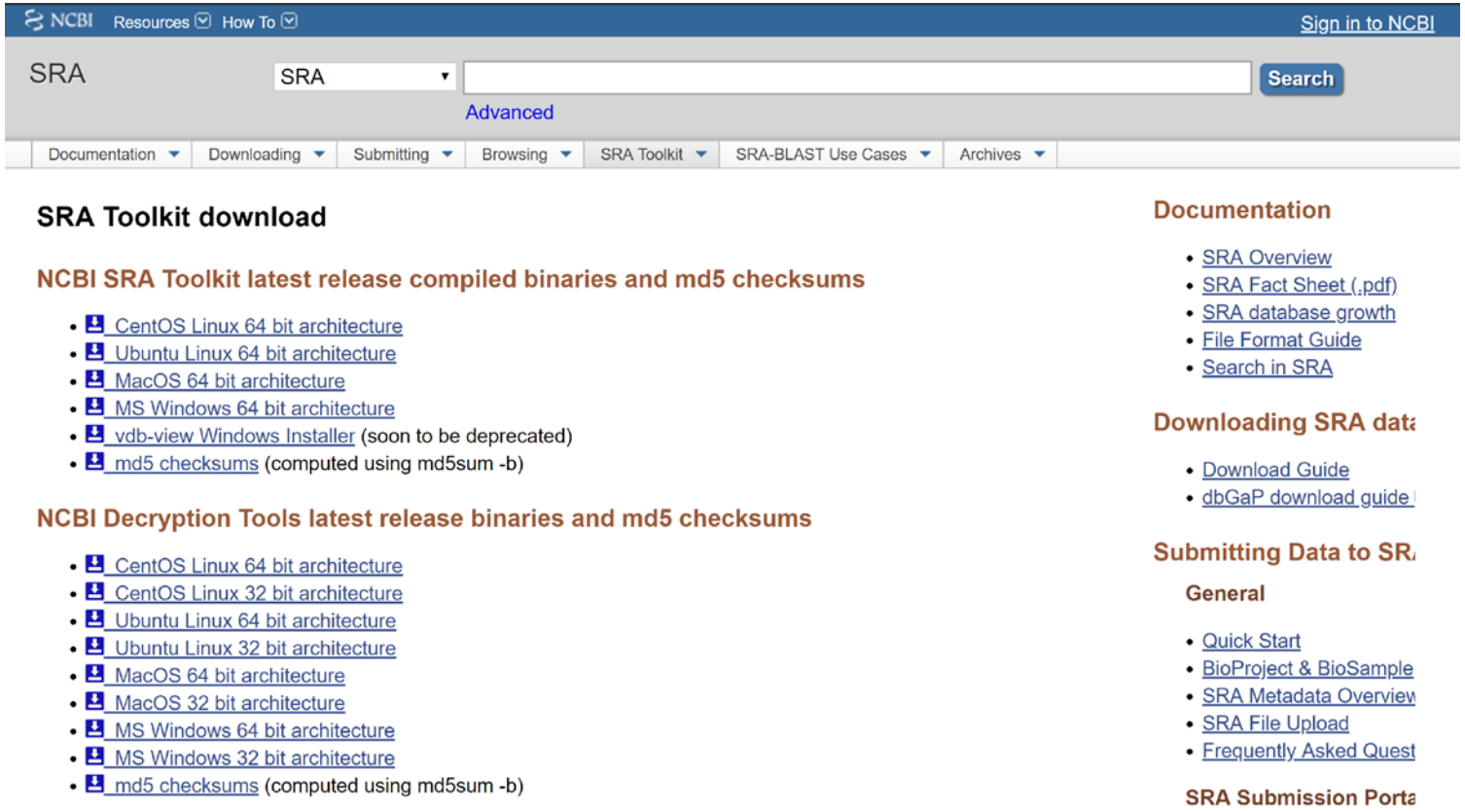
#	Study, Consent	Status	Expiration	Application
	Add Health: The National Longitudinal Study of Adolescent to Adult Health (Add Health) (phs001367.v1.p1) General Research Use (IRB, PUB, GSO) (phs001367.v1.p1.c1), NICHD	New		view

Back | Return to My Projects

NIH Genotype and Phenotype database is a service of NCBI. Please **contact us** with any questions.

National Center for Biotechnology Information | U.S. National Library of Medicine

To Download dbGaP data; Download NIH SRA Toolkit



Navigate to https://www.ncbi.nlm.nih.gov/sra/docs/toolkitsoft/

Download relevant architecture (CentOS or Ubuntu Linux)

tar -xvzf /path/to/file/sratoolkit.current-ubuntu64.tar.gz

Select files on dbGaP

- Navigate to dbGaP https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?page=login and login using the eRA account credentials
- Click on “My Requests” tab. The list of Approved Requests is under “Approved” sub-tab. (slide 24)
- Find the table row of approved dataset (phs001367.v1.p1), click on the link named “Request Files” in the “Actions” column.
- On the “Access Request” page, go to the “Phenotype and Genotype files” sub-tab and click on the “dbGaP File Selector” link.
- Add/remove files using the facets listed in the left panel facet manager. From the right panel file list, select/unselect files by checking/unchecking checkboxes in front of the file names. (slide 25)
- Once the files are selected (checked), click on the “Cart File” button (on the upper part of the page) and save the cart file (.kart).

dbGaP Approved Requests

NCBI Site map | All databases | PubMed | Search

db GaP genotypes and phenotypes

Logged in as **Kathleen Harris** | Log out

Browse/Search | **Authorized Access** | Help

My Submitted Data | Data Browser | My Projects | **My Requests** | Downloaders | My Profile

Request List

#	Study, Consent	Status	Expiration	Actions

NIH Genotype and Phenotype database is a service of NCBI. Please **contact us** with any questions.

National Center for Biotechnology Information | U.S. National Library of Medicine

Privacy Notice | Disclaimer | Accessibility

NIH FIRSTGOV.gov

dbGaP File Selector

NCBI dbGaP File Selector | Help | Permalink | Logged in dbGaP as **Kathleen Harris** | Log out

Facets

- Content type
- File name
- Size
- Data category
- Embargo date
- File accession
- Genotype platform

Hide common fields

Consent:	GRU-IRB-PUB-GSO
Downloaded:	no
Release date:	2018-09-12
Study accession:	phs001367.v1.p1

	Files	Size	Download
Total:	25	637.95 Gb	Files Table
Selected:			Cart File

25 Files found

Content type	File name	Size	Description	Data category	Embargo date	File accession	Genotype platform
Use contents	Study_Report.phs001367.AddHealth.v1.p1.MULTI.pdf	55.2 kb	Master stu...d consent category).	StudyMeta	No Embargo	phs001367.v1.p1	N/A
Use contents	Release_Notes.phs001367.AddHealth.v1.p1.MULTI.pdf	144.1 kb	Release no...wnload component set	StudyMeta	No Embargo	phs001367.v1.p1	N/A
Use contents	manifest_p....GRU-IRB-PUB-GSO.pdf	41.8 kb	Release no...wnload component set	StudyMeta	No Embargo	phs001367.v1.p1	N/A
Phenotype data-dictionary	phs001367....ubject.data_dict.xml	734b	pht008245....ataset pht008245.v1.	Phenotype	2018-09-12	phs001367.v1.p1	N/A
Phenotype individual-auxiliary	phs001367....Subject.MULTI.txt.gz	67.2 kb	pht008245....onsent - Information	Phenotype	2018-09-12	pht008245.v1.p1	N/A
Phenotype variable-report	phs001367....bject.var_report.xml	2.0 kb	pht008245....ataset pht008245.v1.	Phenotype	2018-09-12	pht008245.v1.p1	N/A
Phenotype data-dictionary	phs001367....digree.data_dict.xml	1.1 kb	pht008246....ataset pht008246.v1.	Phenotype	2018-09-12	phs001367.v1.p1	N/A
Phenotype individual-pedigree	phs001367....edigree.MULTI.txt.gz	29.5 kb	pht008246.v1: Pedigree Information	Phenotype	2018-09-12	pht008246.v1.p1	N/A
Phenotype variable-report	phs001367....igree.var_report.xml	4.3 kb	pht008246....ataset pht008246.v1.	Phenotype	2018-09-12	pht008246.v1.p1	N/A
Phenotype data-dictionary	phs001367....Sample.data_dict.xml	569b	pht008247....ataset pht008247.v1.	Phenotype	2018-09-12	phs001367.v1.p1	N/A
Phenotype individual-auxiliary	phs001367...._Sample.MULTI.txt.gz	284.1 kb	pht008247....mple Use information	Phenotype	2018-09-12	pht008247.v1.p1	N/A
Phenotype variable-report	phs001367....ample.var_report.xml	2.2 kb	pht008247....ataset pht008247.v1.	Phenotype	2018-09-12	pht008247.v1.p1	N/A
Phenotype data-dictionary	phs001367....otypes.data_dict.xml	1.4 kb	pht008248....ataset pht008248.v1.	Phenotype	2018-09-12	phs001367.v1.p1	N/A
Phenotype variable-report	phs001367....types.var_report.xml	9.4 kb	pht008248....ataset pht008248.v1.	Phenotype	2018-09-12	pht008248.v1.p1	N/A
Phenotype individual-traits	phs001367....U-IRB-PUB-GSO.txt.gz	170.9 kb	pht008248.... height information.	Phenotype	2018-09-12	pht008248.v1.p1	N/A
Phenotype data-dictionary	phs001367....ibutes.data_dict.xml	930b	pht008249....ataset pht008249.v1.	Phenotype	2018-09-12	phs001367.v1.p1	N/A
Phenotype variable-report	phs001367....butes.var_report.xml	3.7 kb	pht008249....ataset pht008249.v1.	Phenotype	2018-09-12	pht008249.v1.p1	N/A
Phenotype individual-traits	phs001367....U-IRB-PUB-GSO.txt.gz	70.3 kb	pht008249.... source of samples.	Phenotype	2018-09-12	pht008249.v1.p1	N/A
Genotype sample-information	phg001069.v1.AddHealth.sample-info.MULTI.tar.gz	247.2 kb	Informatio...files in the release	Genotype	2018-09-12	phg001069.v1	NULL
Genotype calls-matrix-format	phg001069....U-IRB-PUB-GSO.tar.gz	846.3 Mb	Set of tex...a particular consent	Genotype	2018-09-12	phg001069.v1	NULL
Genotype qc	phg001069...._v1-0_H.MULTI.tar.gz	6.4 Mb	Marker and...requnecy, and others	Genotype	2018-09-12	phg001069.v1	NULL
Genotype imputed-data	phg001099....-PUB-GSO.set1.tar.gz	518.5 Gb	Imputed ge...me and subject group	Genotype	2018-09-12	phg001099.v1	NULL
Genotype imputed-data	phg001099....-PUB-GSO.set2.tar.gz	118.7 Gb	Imputed ge...me and subject group	Genotype	2018-09-12	phg001099.v1	NULL
Genotype sample-information	phg001099.v1.AddHealth.sample-info.MULTI.tar.gz	2.4 Mb	Informatio...files in the release	Genotype	2018-09-12	phg001099.v1	NULL

Create directory for results

- **Naming convention isn’t optional, requires lower case ‘ncbi’**
- Create a directory titled ‘ncbi’ at root
- *cd ~; mkdir ncbi; cd ncbi*
- Move .kart file into ncbi directory ie. *mv cart_prj19687_201902081009 ./*

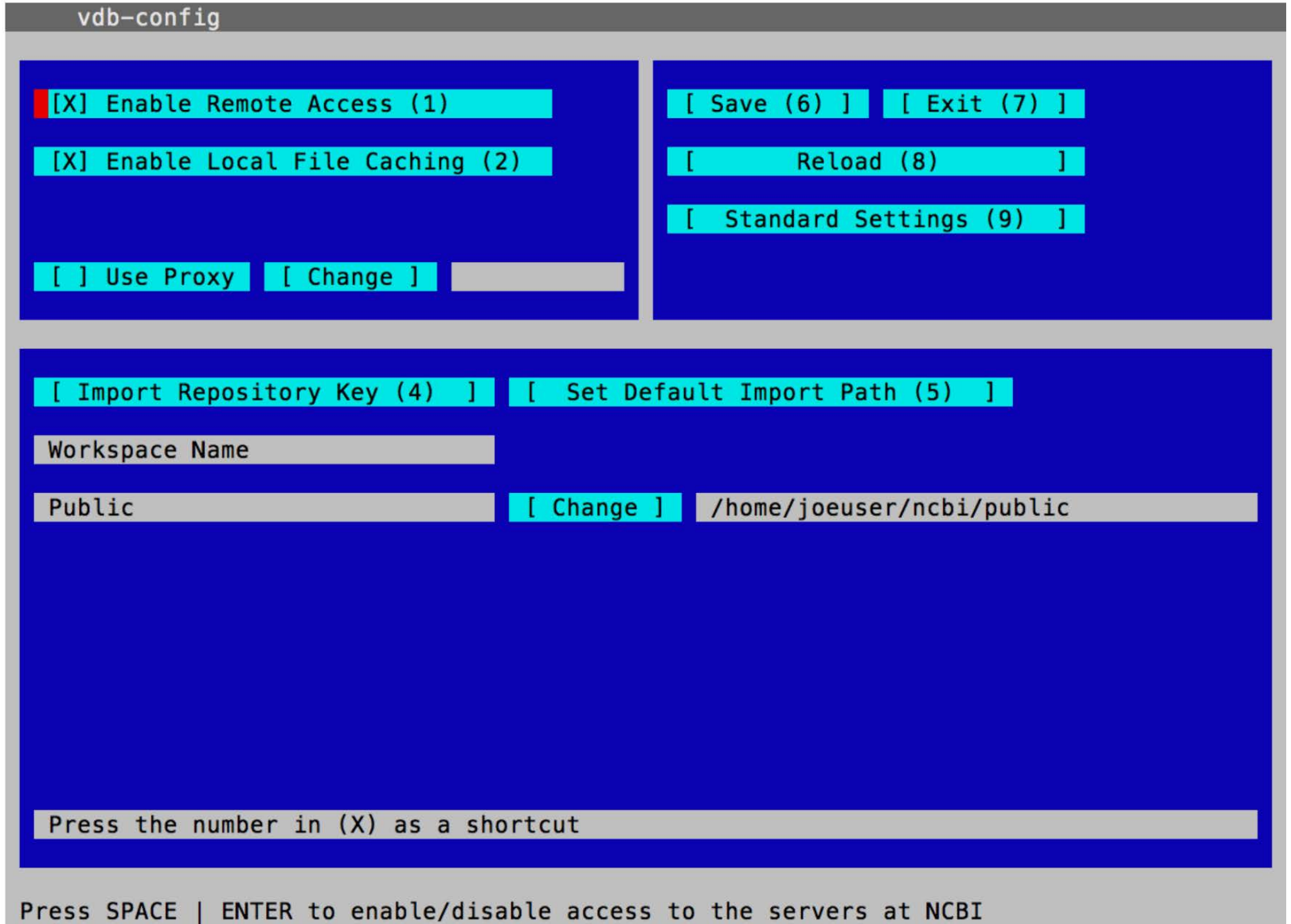
prefetch dbGaP files

- Use the ‘prefetch’ utility to download the data files specified by the cart file.
- */path-to-sratoolkit-install-dir/bin/prefetch -t ascp –a cart_prj19687_201902081009*
- Depending on the size, you might need to specify ‘*–max-size 100000000*’
- This will generate a directory at ~/ncbi/dbGaP-19687 which contains files, sra, wgs, nannot, and refseq
- All data and metadata will be in ~/ncbi/dbGaP-19687/files
- All files will be encrypted and end *.ncbi_enc

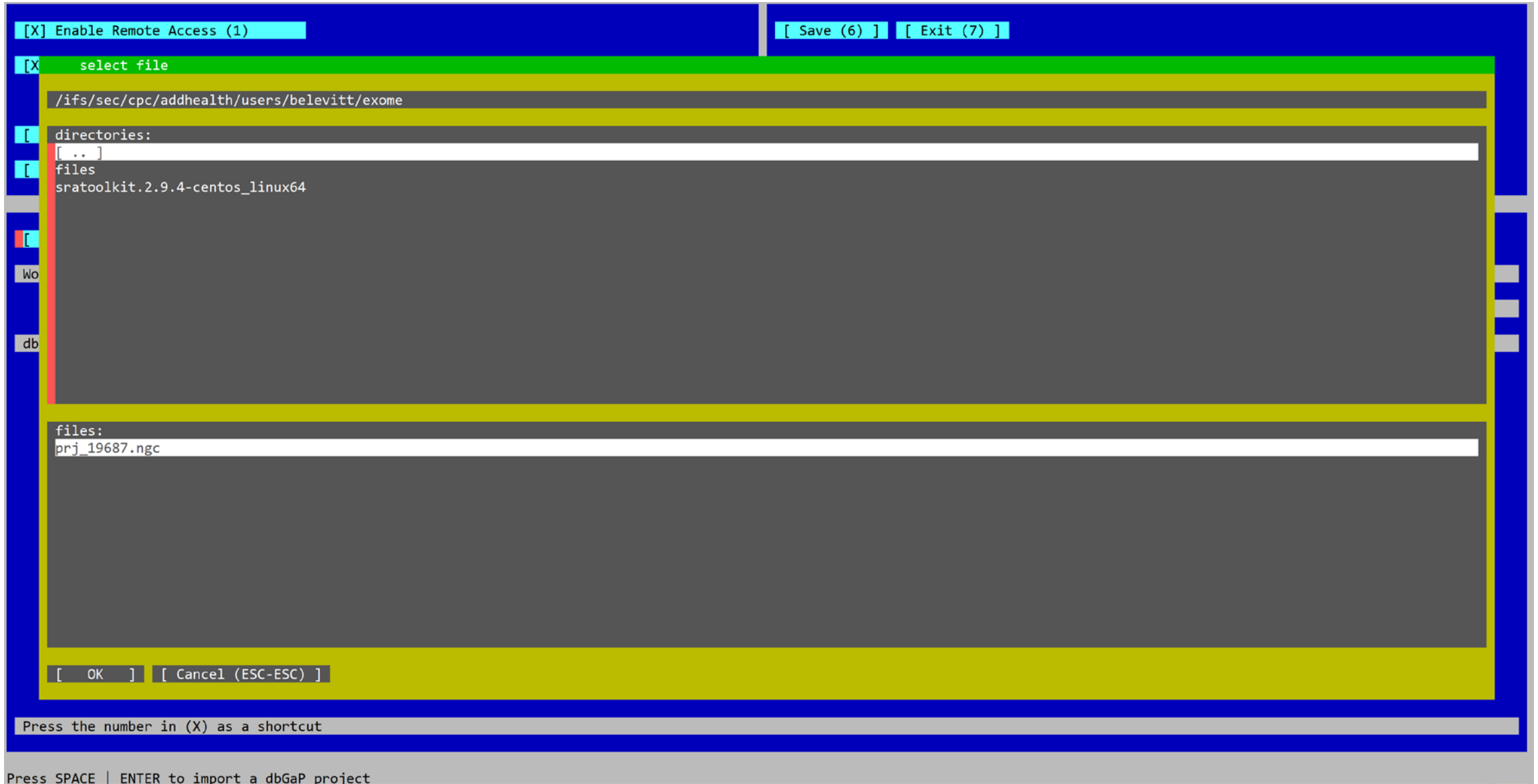
Decrypt files with vdb-config

- Use vdb-config tool to decrypt them
- *ncbi/dbGaP_19687/files$ /path/to/sratoolkit.2.9.4-centos_linux64/bin/vdb-config –i*
- This will open an interactive vdb-config session. (slide 29)
- Import your repository key (NGC file), box 4 (slide 30)
- Select your folder eg. ncbi/dbGaP-19687, tab down to ‘change’ (slide 31)
- Click save, box 6
- Click exit, box 7

vdb-config interface



Assign repository key



Decrypt files

```
vdb-config

[X] Enable Remote Access (1)                                  [ Save (6) ]  [ Exit (7) ]

[X     select directory

       /ifs/sec/cpc/addhealth/users/belevitt/ncbi/dbGaP-19687/files

[      directories:
       [ .. ]
[      QC
       imputed
       matrix
[      phg001099.v1.AddHealth.marker-info.MULTI
       phg001099.v1.AddHealth.sample-info.MULTI
Wo     sample
       tarballs

db

       [   OK   ]  [ Cancel (ESC-ESC) ]  [  Goto  ]  [  Create Dir  ]

Press the number in (X) as a shortcut

Press SPACE | ENTER to change location for this dbGaP repository
```

Additional resources for downloading dbGaP data

https://www.ncbi.nlm.nih.gov/books/NBK36439/ GaP FAQ Archive: Downloading Data
https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?page=login dbGaP authorized access point
https://www.biostars.org/p/316506/ option for prefetch large files
https://github.com/ncbi/sra-tools/wiki/Toolkit-Configuration how to navigate the vdb-config utility

Additional Add Health OMICs Resources

- Sign up for the addhealthomics listserv. To subscribe/join:
  1. Send an email to subscribe-addhealthomics@listserv.unc.edu with no message body
  2. Wait for addhealthomics confirmation email
  3. Click the confirm link the email
- ***Future resources*** that will aid users in accessing, understanding, analyzing, and interpreting Add Health genomics data, prioritizing GWAS data, will be posted at:
  - https://www.biostars.org/t/addhealthomics/
  - https://www.cpc.unc.edu/projects/addhealth/documentation/omics